



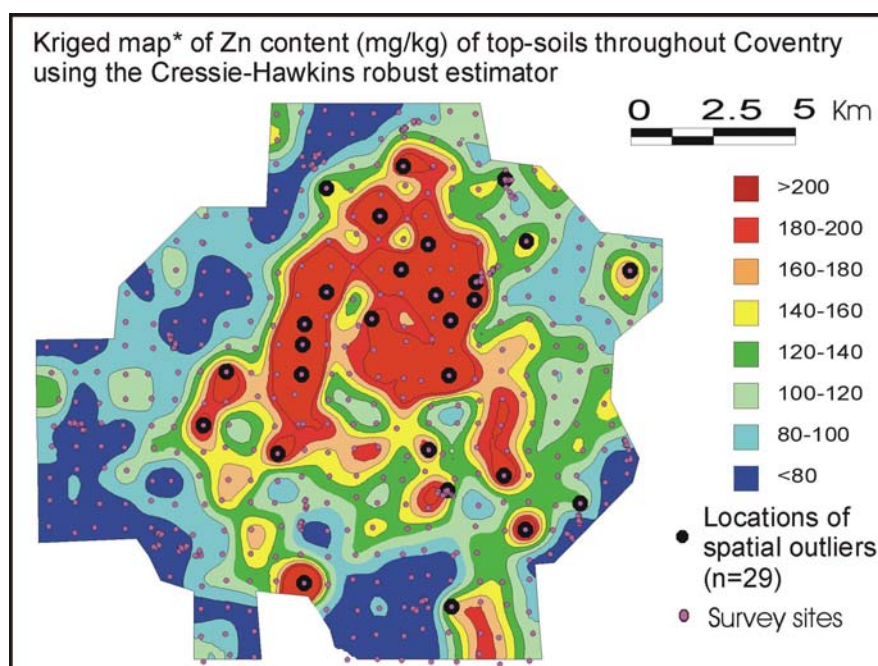
**British  
Geological Survey**

NATURAL ENVIRONMENT RESEARCH COUNCIL

# Assessing geostatistical methods for presenting urban soil geochemical data from Coventry

Urban Geoscience and Geological Hazards Programme

Internal Report IR/03/012





BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/03/012

# Assessing geostatistical methods for presenting urban soil geochemical data from Coventry

Barry Rawlins and Sarah Brown

The National Grid and other  
Ordnance Survey data are used  
with the permission of the  
Controller of Her Majesty's  
Stationery Office.  
Ordnance Survey licence number  
GD 272191/1999

## *Key words*

Geochemistry, robust,  
geostatistics, estimator,  
variogram, kriging, Coventry,  
nested survey.

## *Front cover*

Isarithmic map of the zinc  
content of top-soils throughout  
Coventry.

## *Bibliographical reference*

RAWLINS, B. G. AND BROWN, S.  
E. 2002. Assessing geostatistical  
methods for presenting urban soil  
geochemical data from Coventry  
*British Geological Survey  
Internal Report, IR/03/012.*  
31pp.

## BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham and Edinburgh; see contact details below or shop online at [www.thebgs.co.uk](http://www.thebgs.co.uk)

The London Information Office maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

*The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.*

*The British Geological Survey is a component body of the Natural Environment Research Council.*

### **Keyworth, Nottingham NG12 5GG**

☎ 0115-936 3241 Fax 0115-936 3488  
e-mail: [sales@bgs.ac.uk](mailto:sales@bgs.ac.uk)  
[www.bgs.ac.uk](http://www.bgs.ac.uk)  
Shop online at: [www.thebgs.co.uk](http://www.thebgs.co.uk)

### **Murchison House, West Mains Road, Edinburgh EH9 3LA**

☎ 0131-667 1000 Fax 0131-668 2683  
e-mail: [scotsales@bgs.ac.uk](mailto:scotsales@bgs.ac.uk)

### **London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE**

☎ 020-7589 4090 Fax 020-7584 8270  
☎ 020-7942 5344/45 email: [bgs london@bgs.ac.uk](mailto:bgs london@bgs.ac.uk)

### **Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU**

☎ 01392-445271 Fax 01392-445371

### **Geological Survey of Northern Ireland, 20 College Gardens, Belfast BT9 6BS**

☎ 028-9066 6595 Fax 028-9066 2835

### **Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB**

☎ 01491-838800 Fax 01491-692345

### *Parent Body*

### **Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU**

☎ 01793-411500 Fax 01793-411501  
[www.nerc.ac.uk](http://www.nerc.ac.uk)

# Acknowledgements

A large number of individuals in the sample preparation team and XRFS laboratories at Keyworth, and on the G-BASE team (staff and volunteer workers) have contributed to the project. We would also like to thank Ellie Steadman and Hubert Murphy for the particle-size analysis of the nested survey soil samples, Dr Murray Lark from the Silsoe Research Institute (SRI) for undertaking aspects of the geostatistical analysis, and Dr Richard Webster for helping with the Unbalanced ANOVA analysis.

## Contents

<b>Acknowledgements .....</b>	<b>i</b>
<b>Contents .....</b>	<b>i</b>
<b>Summary .....</b>	<b>iv</b>
<b>1 Introduction and Aims .....</b>	<b>6</b>
1.1 The need for robust geostatisticians .....	6
1.2 Summary procedure for assessing variogram models .....	7
1.3 Summary procedure for identifying spatial outliers .....	7
1.4 Theory of nested sampling .....	8
<b>2 Urban survey and nested sampling in Coventry .....</b>	<b>9</b>
2.1 Location and history .....	9
2.2 Distribution and types of parent material .....	9
2.3 Geochemical survey .....	11
2.4 Nested survey: site selection .....	12
2.5 Nested survey: sampling procedure .....	13
<b>3 Sample preparation and analysis .....</b>	<b>13</b>
3.1 XRFS analysis .....	13
3.2 Soil pH analysis .....	14
3.3 Loss on ignition analysis .....	14
3.4 Particle-size analysis .....	14
<b>4 Soil geochemical and particle-size data (urban and nested surveys) .....</b>	<b>15</b>
4.1 Geochemistry: summary statistics .....	15
4.2 Cadmium contamination of nested survey sites .....	16
4.3 Particle-size data .....	16
<b>5 Nested analysis of variance .....</b>	<b>18</b>
5.1 Components of variance .....	18
<b>6 Geostatistical analysis .....</b>	<b>18</b>
6.1 Data transformation .....	18
6.2 Assessing the performance of variogram estimators .....	19
6.3 Kriging cadmium with Matheron's standard estimator .....	20

6.4 Identification of spatial outliers (zinc data).....	22
6.5 Kriging topsoil zinc with and without spatial outliers .....	22
<b>7 Discussion and Conclusions .....</b>	<b>23</b>
7.1 Use of robust versus classical geostatistics .....	23
7.2 The objectives of undertaking geochemical surveys in the urban environment .....	24
7.3 Sampling intensity versus estimation uncertainty .....	24
7.4 Targeted sampling .....	24
<b>Glossary .....</b>	<b>25</b>
<b>References.....</b>	<b>27</b>

## FIGURES

Figure 1 - Soil sample sites and parent materials in the urban and nested survey (n=551). .....	10
Figure 2 - Sampling configuration used at sites in the urban and nested surveys .....	11
Figure 3 - Spatial configuration of a set of sampling points from one centre in the nested survey .....	12
Figure 4 - Particle-size distribution of soils in the nested survey (grouped by colour) .....	17
Figure 5 - Components of variation by stage of the nested survey .....	17
Figure 6 - Experimental variograms (symbols) and models fitted to them (lines) for log <sub>e</sub> transformed data. ....	19
Figure 7 - Isotropic spherical model (solid line) fitted to the experimental variogram (circles) using Matheron's estimator - log <sub>e</sub> transformed Cd data. ....	20
Figure 8 – Radii of circles around an unsampled location (x) where a value is to be estimated in relation to actual sample points (●) on a 500 metre* sampling interval. ....	21
Figure 9 - Kriged map of Cd content (mg kg <sup>-1</sup> ) of top-soils throughout Coventry using Matheron's standard estimator .....	21
Figure 10 - Isotropic spherical model (solid line) fitted to experimental variogram (circles) using the Cressie-Hawkins robust estimator: log <sub>e</sub> transformed Zn data. ....	22
Figure 11 - Kriged maps showing Zn content (mg kg <sup>-1</sup> ) of top-soils in Coventry : a) with, and b) without, spatial outliers .....	23

## TABLES

Table 1 - Cumulative samples at each level in the nested sampling design .....	8
Table 2 - Derivation of components of variance for an unbalanced design .....	9
Table 3 – Proportion of land use types in Coventry recorded during the urban survey .....	9
Table 4 - Locations and land use at the thirteen nested sites .....	12
Table 5 Lower (LLD) and upper reporting limit (URL) values for XRFS analysis of G-BASE urban soil samples .....	13
Table 6 - Summary statistics for combined urban and nested survey of top-soil geochemistry in Coventry (n=551) .....	15
Table 7 - Average clay, silt and sand content of soils at each of the nested survey centres .....	16

Table 8 - Descriptive statistics for the total concentration of five elements in urban and nested soil surveys (n=551) used in the geostatistical analysis and the skewness co-efficients of the $\log_e$ transformed data. ....	18
Table 9 – Parameters of the variogram models and the cross validation assessment. ....	19

## Summary

The current shortest distance between sample points in urban soil surveys of the BGS GSUE (Geochemical Surveys in Urban Environments) project is around 350 metres. Geostatistical analysis using soil geochemical data from the survey of Stoke-on-Trent had shown that much of the variation occurs at shorter sampling intervals (less than 350 metres). This means that the uncertainties associated with estimating values at unsampled locations (using kriging) are likely to be relatively large. A project was designed to address three specific, and related, questions. First, what is the nature of the short-scale variability of major and trace elements in urban environments? Second, would recently published, robust geostatistical methods be more appropriate for producing interpolated maps of urban soil geochemistry in which there may be two processes operating; a background process and a quasi-point contaminant process? Third, is the current sampling resolution (adopted in urban surveys) appropriate?

To address the first question we undertook a 'nested' survey at selected nodes within the urban area of Coventry where a standard GSUE survey had recently been undertaken. These samples were analysed for the same suite of major and trace elements as the standard survey, and also for their particle-size distribution (proportions of sand, silt and clay). **In the case of the typical contaminant type elements (Pb, Sn, Sb, Cd, Ni, Cu and Zn) a large proportion (22-75%) of the variation occurred at spatial scales of 30 metres or less, compared to a range of major and trace elements and particle size classes.** Without further chemical analysis, it is not possible to determine whether this variation is due to anthropogenic impacts (pollution) or natural variation. However, many of these elements are common environmental pollutants suggesting that their greater short-scale variability may result from human activities. **The information on variability at short spatial scales collected in the nested survey enabled us to plot variograms for a range of elements in which almost all the spatial variation was captured.**

Robust geostatistical methods may be more appropriate when dealing with datasets in which there are a considerable number of outlying values. An assessment of conventional and robust geostatistical estimators was undertaken based on five elements with significantly skewed distributions (Cd, As, Pb, Zn and Ni). Based on the results of a cross-validation exercise, the conventional geostatistical estimator (that due to Matheron) was found to be optimal for estimating Cd, As, Pb and Ni. However, in the case of the data for Zn, in which there were a considerable number of outlying values, a robust estimator (Cressie-Hawkins) performed best. **If optimal interpolation methods are to be used in mapping urban soil geochemistry, it is recommended that when a large number of outliers are present in a dataset, a comparison of robust and classical estimators is undertaken.**

In the case of Zn, another statistical technique was used to identify 29 spatial outliers - samples which appear to be the outcome of a quasi-point process. Interpolated maps were generated both with and without the spatial outliers to determine the scale of their impact on the background process. In the case of the latter, there was a significant difference in the distribution in the region of the highest values. In addition, a number of the 'bullseye' patterns which were associated with spatial outliers have been removed. **This is a useful technique for separating the background process from the effects of a quasi-point process. However, it is computationally demanding and time-consuming. It remains to be seen whether a customer from a local authority would be prepared to pay for this level of skilled analysis in the preparation of contour maps. To determine whether robust geostatistical methods should be applied to other urban data, there is a need to determine the skewness coefficients and the number of outliers in datasets from the other urban centres. This will provide a better understanding of whether the Coventry data is typical of the other centres for which data are available.**

Geostatistical analysis of urban geochemical data to date has indicated that the current sampling resolution adopted in urban surveys does capture a varying proportion of the spatially correlated



variation for most of the contaminant type elements. **On the basis of the work presented here we would not advocate changing the current (4 samples per square kilometre) sampling resolution.** However, when producing a continuous surface map, it is essential first to undertake exploratory data analysis and construct variograms of the data to assess the degree of autocorrelation prior to kriging. If little or no autocorrelation is present, it is preferable to present the data as proportional symbols because there is little or nothing to be gained from interpolation. When a continuous surface map has been produced it should be accompanied by a description of the method used to create it, and a statement that the contour intervals represent estimates, not true values.

The nested survey also identified Cd (cadmium) contamination at the site of a series of allotments. Comparison with recently published soil guideline values suggests that it may represent a potentially significant risk to human health. This issue is currently being raised with the City Council. **To increase the utility of the soil geochemical data collected in the urban environment, we recommend that it would be beneficial to focus on perhaps two or three key sites in an urban area where human exposure to contamination may be significant, such as allotments, children's nurseries or groundwater protection zones.** These sites could be selected in conjunction with the City/Local Council on the basis of land use information which may indicate the likely presence of historical contamination.

# 1 Introduction and Aims

Under its Geochemical Survey in Urban Environments (GSUE) project, BGS has collected and analysed soil samples for their total concentration of major and trace elements in around 20 urban centres. To date, these data have been presented and licensed to local authorities as proportional symbol maps, as opposed to continuous (interpolated) surfaces mainly because of the possibility, in using the latter method, of overestimating the extent of elevated concentrations around hot-spots of contamination which have been sampled.

In the rural environment, variation in soil geochemistry shows very clear autocorrelation<sup>1</sup> (samples close together are more similar than those farther apart). Much of the variation is captured when sampling at spatial scales of 1 sample in every other kilometre square of the British National Grid – the shortest distance between samples being around 700 metres. In the urban environment, the variation occurs at shorter spatial scales and there are numerous sites with outlying values that do not conform to the general pattern of variation. Previous work (Fordyce, 2003) had indicated that much of the spatial variation occurs at distances shorter than 7350 metres (the shortest distance between sample sites in the GSUE urban survey). To increase our understanding of the spatial variability at shorter scales some further sampling was undertaken in Coventry in which samples were separated by shorter lag intervals (known as a nested survey (Oliver and Webster, 1986). By creating variograms for specific chemical elements in an urban centre, it is possible to calculate the sampling resolution required to give a maximum estimation error based on a regular sample grid. This can be used to design optimal sampling schemes based on a required estimation uncertainty for unsampled locations.

Statistical analysis of soil geochemical datasets in which a significant proportion of sample sites have elevated values of trace metal contents has shown that the use of robust geostatistical methods may be more appropriate than classical geostatistics (Lark, 2002).

The aims of this study were to:

- determine the nature of spatial variability of major and trace elements at a range of spatial scales across one urban centre (by calculating empirical variograms based on the urban and nested survey data)
- test both classical and robust geostatistical methods to determine which are more appropriate when mapping (using kriging) urban soil geochemistry, and identify the key factors which determine this
- make recommendations concerning the presentation of urban soil geochemical data based on the typical sampling resolution (4 samples per square kilometre) – both for BGS reports and city councils which may licence the data

## 1.1 THE NEED FOR ROBUST GEOSTATISTICS

In the urban environment, the background variation of a soil property may not resemble a Gaussian random variable, but the combination of two superimposed processes: continuous

---

<sup>1</sup> Much of the language and theory underlying this report assumes the reader has some familiarity with the topic of geostatistics. This is a branch of statistics concerning the sampling and estimation of natural, spatially varying phenomena (for example the concentration of lead in the soil across a county or region). Good introductory texts on this topic include *Burrough, P A, and McDonnell, R A. 2000. Principles of Geographical Information Systems. (Oxford University Press)* and Webster, R, and Oliver, M A. 2001. *Geostatistics for Environmental Scientists* (Chichester: John Wiley & Sons).. Many of the geostatistical terms, which may not be familiar to the reader, are defined in the glossary at the end of the report.

background variation and a quasi (contaminant) point process. It has been shown that the quasi point process will have a large effect on estimates of the variogram (Lark, 2000) and that kriging from these data may produce a surface which is strongly influenced by outlying values. It would also be useful to separate these two processes (the background from the quasi point process) so that each could be dealt with separately. Lark (2002) proposed a formal method for assessing variogram models and, where the use of a robust estimator was appropriate, identifying spatial outliers (from the contaminant process). The method for assessing variogram models and identifying spatial outliers is summarised below.

## 1.2 SUMMARY PROCEDURE FOR ASSESSING VARIOGRAM MODELS

The full details of the methods are described by Lark (2002)

- Conduct exploratory analysis on the variate (e.g. lead concentration) to determine whether transformation is necessary. If the background process is not Gaussian then transformation is needed. Two statistics can be used to assess the distribution: the skew and octile skew coefficients.

The octile skew is defined by the following formula, where P stands for percentile<sub>(x)</sub>, where (x) can take values from 0 to 1:

$$OS = \frac{(P_{0.875} - P_{0.5}) - (P_{0.5} - P_{0.125})}{(P_{0.875} - P_{0.125})}$$

If the standard skewness co-efficient<sup>2</sup> is larger than 1 (or less than -1), this suggests that the distribution may not be Gaussian. However, if the value of the octile skew is less than 0.2 (or greater than -0.2), then it may not be appropriate to transform the data which are likely to have been influenced by a few outlying values.

- Transform the data where necessary. For example, transform all values by taking natural logarithms and then check the skew is between -1 and +1 (octile skew in the range -0.2 to +0.2).
- Estimate variograms of the data using Matheron's standard estimator and robust estimators: Cressie-Hawkins (Cressie and Hawkins, 1980) and Dowd (1984).
- Estimate the standardized kriging error and its square by cross validation and obtain the median for each variogram. If the median value using Matheron's estimator is not significantly different from 0.455, this suggests there are no spatial outliers in the data, and a conventional geostatistical analysis may be performed.
- If the median value of the squared standardised kriging error (using Matheron's estimator) is significantly different from 0.455, then the robust estimator giving the closest value should be selected as the most appropriate model. The spatial outliers in the dataset can then be identified using the method described in the next section.

## 1.3 SUMMARY PROCEDURE FOR IDENTIFYING SPATIAL OUTLIERS

- The standardized error calculated by cross-validation<sup>3</sup>:

---

<sup>2</sup> Skewness measures the deviation of the distribution from symmetry. If the skewness is clearly different from 0, then that distribution is asymmetrical, while normal distributions are perfectly symmetrical.

$$\text{Skewness} = n * M_3 / [(n-1) * (n-2) * \sigma^3]$$

Where  $M_3$  is equal to:  $\sum (x_i - \text{Mean}_x)^3$ ,  $\sigma^3$  is the standard deviation (sigma) raised to the third power and n is the valid number of cases.

<sup>3</sup> Symbols used here are defined in the glossary of terms

$$\varepsilon(x_i) = \frac{\{z(x_i) - \hat{Z}(x_i)\}}{\hat{\sigma}_K(x_i)}$$

will tend to be large and positive at a location where the measured datum is an outlier in the upper tail of the distribution, and large and negative for an outlier from the lower tail. (Lark, 2002) proposed combining the standardized error with the standardized value of the datum  $z_S(x_i)$ :

$$z_S(x_i) = \frac{z(x_i)}{\hat{S}}$$

where  $\hat{S}$  is a robust measure of scale. At each location (x) compute the correlation of the kriging error  $z(x_i) - \hat{Z}(x_i)$  with the true value  $z(x_i)$ , and then compute the projection of the vector  $(z_S(x_i), \varepsilon(x_i))$ , onto the principal axis of their joint distribution function. This test statistic,  $O(x)$ , is a standard normal variable of mean zero under a null hypothesis that the datum  $z(x_i)$  belongs to the underlying normal process and not the contaminating process.

If  $O(x)$  is significantly larger than zero, this provides evidence for regarding the datum value  $z(x_i)$  as representing the contaminant process.

#### 1.4 THEORY OF NESTED SAMPLING

Previous analysis of urban soil geochemical data had indicated that much of the variation occurred at spatial scales shorter than the minimum sampling interval (Fordyce, 2003). A cost-effective way to measure short-scale variability is to undertake nested sampling, as described by Oliver and Webster (1986). The theory is based on the notion that the population of interest can be divided up into distinct stages; classes created by an initial division can be subdivided at a second stage and can be further subdivided in a hierarchical manner. Each stage constitutes a category, and any one sampling unit belongs to one class, and only one class in each category. The idea is that an individual observation embodies a contribution from each stage, including an unresolved variance in the smallest subdivision. The components of variance at each stage can be estimated by designing a sampling scheme in which at any one stage all the classes are divided into the same number of subdivisions.

The nested survey design adopted in Coventry draws upon the method published by Oliver and Webster (1986) who cite the original source (Youden and Mehlich, 1937). The original soil survey of Coventry has 395 sample locations, of which fifty-nine sample pairs have separations of less than 400m. It is therefore necessary to focus sampling effort on lag distances of less than about 300m. The constraints on the nested design were to limit the total number of samples to around 150. By replicating at only half of the third stage units (i.e. collecting four samples at lag distances of 10m for each node of the survey, rather than doubling to eight samples) it is possible to constrain the total number of samples collected to a total of 156 (see Table 1).

**Table 1 - Cumulative samples at each level in the nested sampling design**

Stage	Node number in Figure 3	Interval (m)	Number of samples at this level	Total number of samples (cumulative)
	1		13	13
1	2	270	13	26
2	3	90	26	52
3	4	30	52	104
4	5	10	52	156

The analytical data from the nested survey can be used in two ways. First, to calculate the components of variance at the different spatial scales (10, 30, 90 and 270 metres) for each element / size fraction. The method for deriving the components of variance from the unbalanced design is shown in Table 2. Second, to provide estimates for the experimental variogram at spatial scales shorter than those of the urban survey.

**Table 2 - Derivation of components of variance for an unbalanced design**

Source	Degrees of freedom	Parameters estimated by mean squares
Stage 1	$f_1$	$\sigma_m^2 + u_{1,m-1} \sigma_{m-1}^2 + \dots + u_{1,2} \sigma_3^2 + u_{1,2} \sigma_2^2 + u_{1,1} \sigma_1^2$
Stage 2	$f_2 - f_1$	$\sigma_m^2 + u_{2,m-1} \sigma_{m-1}^2 + \dots + u_{1,2} \sigma_3^2 + u_{2,3} \sigma_3^2 + u_{2,2} \sigma_2^2$
Stage 3	$f_3 - f_2$	$\sigma_m^2 + u_{3,m-1} \sigma_{m-1}^2 + \dots + u_{3,3} \sigma_3^2$
.	.	.
Stage m-1	$f_{m-1} - f_{m-2}$	$\sigma_m^2 + u_{m-1,m-1} \sigma_{m-1}^2$
Stage m	$N - f_{m-1}$	$\sigma_m^2$
Total	$N-1$	

$N$  = sample size

$f_i$  = number of classes at the  $i$ th stage

$u_{ij}$  =  $j$ th coefficient of the variance component at the  $i$ th stage

$\sigma_i^2$  component of variance at the  $i$ th stage

## 2 Urban survey and nested sampling in Coventry

### 2.1 LOCATION AND HISTORY

Coventry is located close to the geographical centre of England. It is a busy manufacturing city, hosting sites for the Jaguar and Peugeot motor companies, with a population of around 300,000. The M6, A45 and A46 trunk roads traverse the area providing valuable links between Coventry and Birmingham. Much of the city was devastated during the World War II blitz (1940 – 1941); it was a prime target as important munitions factories were sited there. Basic observational field data were recorded during the G-SUE survey, including current land use information summarised in Table 3.

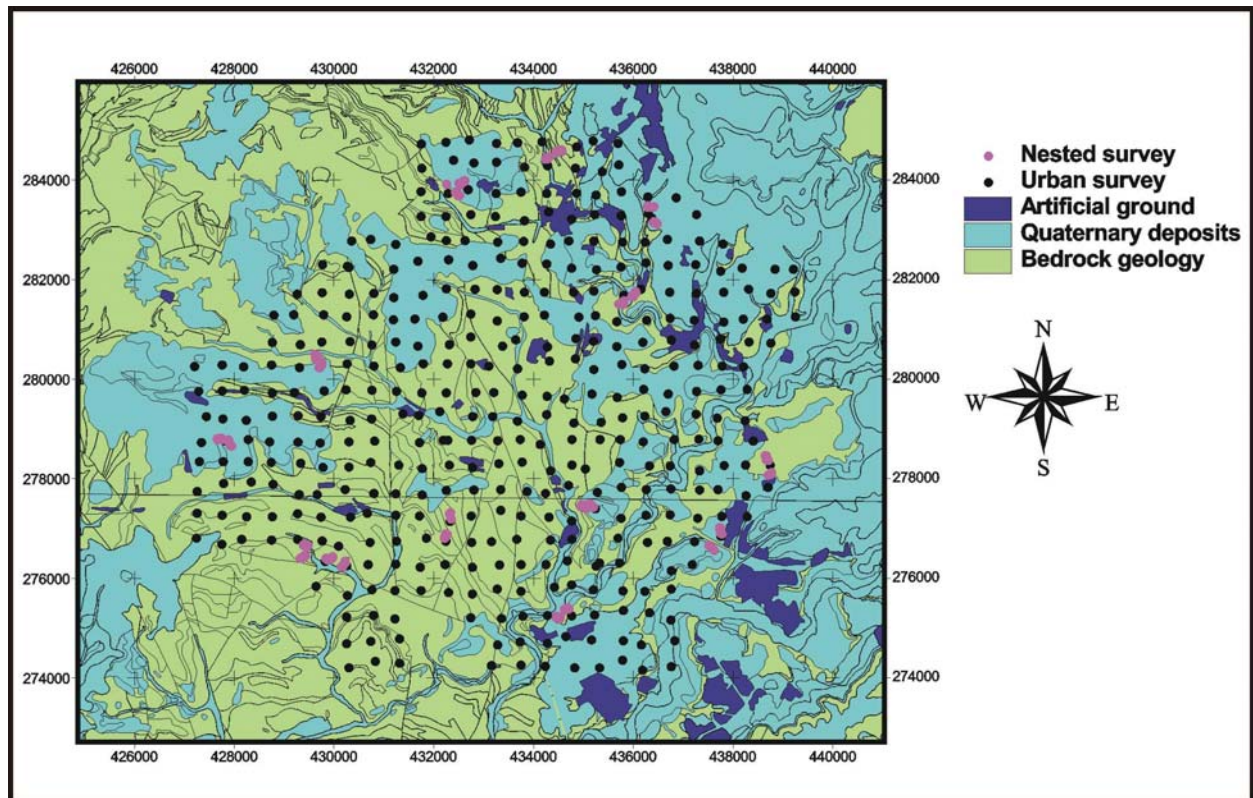
**Table 3 – Proportion of land use types in Coventry recorded during the urban survey**

Land use type	Proportion (%)
Urban (built up areas, housing, shops etc...)	50
Recreational (parks, school playing fields)	24
Farming (arable and livestock)	18
Industrial	4
Other	4

### 2.2 Distribution and types of parent material

Parent material refers to the deposit from which the overlying soil was formed, and can take the form of bedrock and a range of Quaternary or artificial deposits. Alternatively, soil at a site may

comprise what is commonly referred to as made ground, having been put in place through development activities. The distribution of the three main types of parent material is shown in Figure 1, and a summary of the nature of the material is provided below, based on published and unpublished 1:50k digital versions of bedrock and Quaternary maps of the region.



**Figure 1 - Soil sample sites and parent materials in the urban and nested survey (n=551).** Grid references are in metres of the British National Grid.

### 2.2.1 Bedrock geology

The geological sequence of bedrock formations in the Coventry area ranges in age from the Cambrian to the Jurassic, although the study area consists solely of Carboniferous and Triassic formations.

The oldest rocks comprise Carboniferous red beds of the Meriden Formation (mudstone and sandstone with subordinate conglomerate and limestone). These occur to the west of the city centre trending in a north-easterly direction, and also running south-east to north-west through central Coventry. The south-west of the city is dominated by the overlying Tile Hill Mudstone Formation composed of red-brown mudstones with subordinate sandstones and siltstones.

Triassic rocks of the Sherwood Sandstone and Mercia Mudstone Groups occur in the eastern part of the city; the former mainly comprises sandstone lithologies, whereas the latter is mudstone dominated.

### 2.2.2 Quaternary deposits

Glacial deposits – there are large areas of till to the west, in the centre and to the north-east of Coventry. Glacio-lacustrine deposits consisting of clay and silt lie to the west and north-west of the city. Sand and gravel outwash deposits occur in discrete areas predominantly to the east of the city.

Post Glacial Deposits – consisting mainly of flights of river terraces bordering the River Avon and its principal tributary. They are composed of sand and gravel, running through the city centre trending south-east to north-west.

### 2.2.3 Artificial Ground

The majority of the artificial ground in Coventry lies to the east of the city centre. The few areas of made ground to the west are likely to influence soil geochemistry in the Canley, Lower Eastern Green and East Hockley areas. To the east there are two major areas of made ground in the Longford and Henley Green areas which have been sampled. There is also a landfill site at Great Heath to the north-west of the city centre, from which overlying soil was sampled.

## 2.3 GEOCHEMICAL SURVEY

An area of 97 km<sup>2</sup> was surveyed during the summer of 2000, in which a total of 389 sites were sampled. This extends from SP2774 (south-west corner) to SP3985 (north-east corner) on the British National Grid, including the city centre and its immediate surroundings, but excluding the built-up areas of Bedworth and Kenilworth.

In the original urban survey, sample sites were arranged on a non-aligned grid pattern at a density of 4 samples per km<sup>2</sup>. Sample spacing was kept as regular as possible, around 500m apart, but was constrained by the ground conditions (e.g. buildings and other constructions). Soil samples were collected from the closest area of open ground to the allocated sample point. Care was taken to avoid contamination from roads, buildings, fences, pylon lines, etc. Ideally samples were collected at least 50 m from any of these obstructions. At each sampling location, five individual samples were collected across a depth range of 0-15 cm using a Dutch auger with a flight diameter of 35 mm, at the corners and centre of a square of side length 20 m (see Figure 2). A composite sample comprising equal quantities from each of the five samples was placed in a Kraft paper bag, labelled and stored for transport to the analytical laboratory.

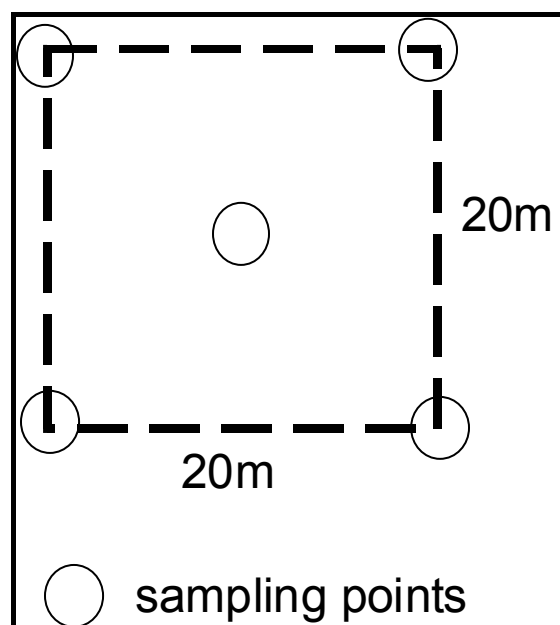


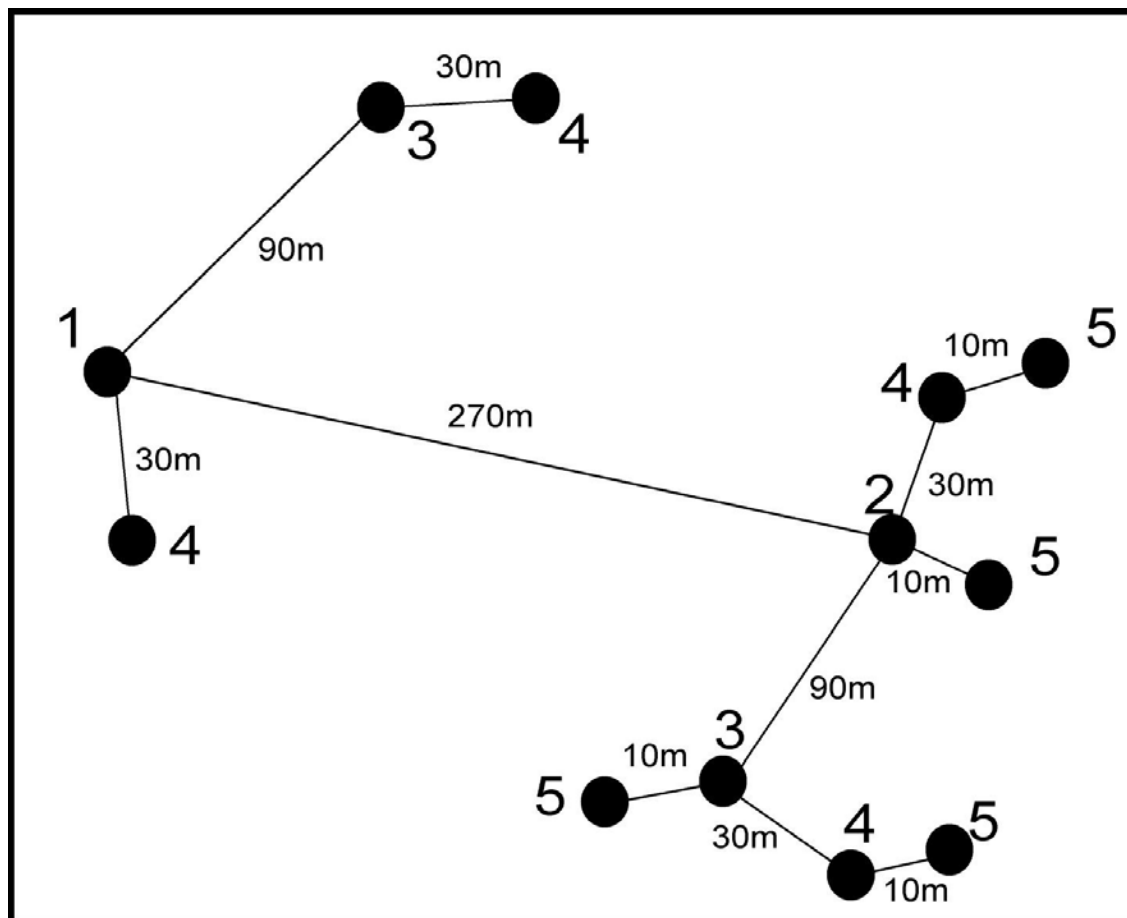
Figure 2 - Sampling configuration used at sites in the urban and nested surveys

## 2.4 NESTED SURVEY: SITE SELECTION

The areas where centres (Figure 3) in the nested survey could actually be used for sampling (due to the constraint of physical space) were limited to areas of open ground (ideally 400 metres by 400 metres). To limit the amount of bias resulting from site selection, our approach was to divide the urban area into thirteen equal squares and randomly select a sufficiently large area of open ground in each one.

**Table 4 - Locations and land use at the thirteen nested sites**

Easting (km)	Northing (km)	Location and land use
434	284	Farm at Longford , Pasture
435	277	London Road, Allotments
432	277	War Memorial Park, Recreational
436	283	Aldermans Green, Nature Reserve
437	276	Wilenhall, Playing field / Waste ground
432	283	Kennedy school, Playing field
427	278	Tilehill Wood School, Playing field
434	275	Roman Fort, Pasture
429	276	University, Playing field
429	276	University, Playing field
429	280	Allesley Hall, Park
435	281	Bell Green, Waste ground
438	278	Binley, Rugby ground/Roadside



**Figure 3 - Spatial configuration of a set of sampling points from one centre in the nested survey**



## 2.5 NESTED SURVEY: SAMPLING PROCEDURE

At each nested survey site, the initial sampling point (denoted 1 in Figure 3) was selected towards the centre of the area of open ground. The same sample procedure was adopted at each sampling point in the nested survey as in the urban survey (see section 2.3). The location of the next sample point was selected by taking the next angle from a random number list (ranging from 0-360 degrees), selecting the direction based on this angle and measuring the distance from the centre of the previous sample site based on the sampling configuration (Figure 3). Where, due to the constraints of space, the random angle resulted in a site location which could not be sampled (e.g. covered ground / buildings), subsequent angles were selected until an appropriate site could be sampled.

## 3 Sample preparation and analysis

### 3.1 XRFS ANALYSIS

All samples were analysed at the BGS laboratories for a range of elements by Wavelength Dispersive X-ray Fluorescence Spectrometry (Ingham and Vrebis, 1994). Three sequential XRF spectrometers were used. A Philips PW1480 fitted with a 216 position sample changer and a 3 kW/100kV tungsten anode X-ray tube was used to determine Cd, Sn and Sb. Two Philips PW2400 spectrometers fitted with 102 position sample changers and with 3 kW/60 kV rhodium anode x-ray tubes were used to determine TiO<sub>2</sub>, MnO, Fe<sub>2</sub>O<sub>3</sub>, V, Cr, Co, and Ba in one suite and Ni, Cu, Zn, As, Mo, Pb, and U in another.

The elements determined and the lower limits of detection (LLD) and upper and lower reporting limits (URL and LRL) for each analyte are shown in Table 5.

The quoted LLDs are theoretical values for the concentration equivalent to three standard deviations above the background count rate for the analyte in a pure silica matrix. High instrumental stability results in practical values for these materials approaching the theoretical.

**Table 5 Lower (LLD) and upper reporting limit (URL) values for XRFS analysis of G-BASE urban soil samples**

Analyte	LLD (ppm)	LRL (%)	URL (ppm)	URL (%)
TiO <sub>2</sub> *	-	0.010	-	100.0
MnO	-	0.010	-	10.0
Fe <sub>2</sub> O <sub>3</sub>	-	0.01	-	100.0
V	2	-	20000	-
Cr	3	-	250000	-
Co	2	-	10000	-
Ni	1	-	4000	-
Cu	1	-	6500	-
Zn	1	-	10000	-
As	1	-	10000	-
Mo	1	-	1000	-
Cd	0.2	-	500	-
Sn	1	-	10000	-
Sb	1	-	10000	-
Ba	3	-	600000	-
Pb	1	-	10000	-
U	1	-	650	-

\* A horizon only.

### 3.2 SOIL pH ANALYSIS

Urban soils were analysed for pH by a slurry method with CaCl<sub>2</sub>, using an epoxy flat tipped pH electrode connected to an Orion Model SA 720 pH meter. The meter was calibrated using buffers pH 4, 7 and 9 and analysis was carried out using 10g of the sample to 25 ml 0.01M CaCl<sub>2</sub>.

### 3.3 LOSS ON IGNITION ANALYSIS

Loss on ignition was carried out by heating each sample to 450°C for a minimum of 4 hours to ensure removal of organic material. It was performed on soil that had been dried overnight at 105°C. The total loss is calculated using the following equation and is expressed as % weight.

$$\left( \frac{\text{weight of crucible \& sample before heating} - \text{weight of crucible \& sample after heating}}{\text{weight of crucible \& sample before heating} - \text{weight of crucible}} \right) \times 100$$

### 3.4 PARTICLE-SIZE ANALYSIS

*Sample Preparation:* The sample was split in half by cone and quartering. Half was retained as reference and the other half was dried at 55°C and weighed for PSD (particle-size distribution) analysis. This was transferred to a screw top leak proof bottle and water added to disaggregate the silt and clay particles then placed on a mechanical shaker for a minimum of 4 hours.

*Wet Sieving:* The silt and clay fractions were removed by wet sieving through a 2 mm sieve nested on a 63 µm sieve and bucket. The sample was washed with water until the water passing the 63 µm sieve was clear. The material retained on the sieves (sand and gravel) was transferred to an evaporating dish and dried at 55°C.

*Dry sieving:* The sand and gravel fractions were dry screened on 2 mm and 63 µm sieves using a mechanical shaker set to run for 15 minutes. The material retained on each size fraction was weighed and the value recorded.

*Sedigraph analysis of the <63 µm fraction:* The < 63 µm suspension from the wet sieving was transferred to 2 litre beakers and dried at 55°C until approximately one litre was remaining. This was then sub split using a peristaltic pump whilst the suspension was being stirred to give approximately 5g of sample for sedigraph analysis. This sub-split was dried then 50 ml of 0.05% calgon solution was added.

X-Ray sedigraph analysis was carried out using a Micromeritics 5100D series analyser. A standard reference material was run before the samples to ensure the machine was operating as it should be. Each sample was analysed on the sedigraph from 0.063 mm to 0.001 mm with cumulative percentage less than data collected at intervals.

Particle-size was calculated from these data as a frequency distribution (i.e. mass retained between each sieve or sedigraph size) and as a cumulative distribution (i.e. mass percentage finer than each sieve or sedigraph size).

## 4 Soil geochemical and particle-size data (urban and nested surveys)

### 4.1 GEOCHEMISTRY: SUMMARY STATISTICS

Prior to further geostatistical analysis, combined summary statistics were calculated for samples collected in the urban and nested surveys (Table 6). The skew statistics indicated that some elements (Na<sub>2</sub>O, Al<sub>2</sub>O<sub>3</sub>, SiO<sub>2</sub>, K<sub>2</sub>O, TiO<sub>2</sub>, MnO, Fe<sub>2</sub>O<sub>3</sub>, U) and pH did not have significantly skewed distributions (skew < 1 or octile skew <0.2), whilst the remainder exhibit varying degrees of skew (skew>1 or octile skew >0.2).

**Table 6 - Summary statistics for combined urban and nested survey of top-soil geochemistry in Coventry (n=551)**

**Values shown are total concentrations in mg kg<sup>-1</sup> (unless otherwise stated)**

	Mean	Median	Min	Max	St. Dev.	Skew	Octile Skew
Na <sub>2</sub> O (%)	0.38	0.40	0.20	0.70	0.08	0.07	0.00
MgO (%)	1.36	1.20	0.30	4.90	0.59	2.51	0.33
Al <sub>2</sub> O <sub>3</sub> (%)	10.4	10.4	4.3	16.0	1.90	-0.02	-0.02
SiO <sub>2</sub> (%)	67.8	68.4	35.5	84.0	5.47	-0.84	-0.16
P <sub>2</sub> O <sub>5</sub> (%)	0.23	0.20	0.07	0.92	0.12	2.39	0.37
K <sub>2</sub> O (%)	2.23	2.14	1.06	4.47	0.54	0.92	0.16
CaO (%)	0.65	0.46	0.06	14.5	0.81	10.1	0.47
TiO <sub>2</sub> (%)	0.53	0.54	0.19	0.84	0.10	-0.34	-0.06
MnO (%)	0.08	0.08	0.01	0.49	0.04	3.50	0.07
Fe <sub>2</sub> O <sub>3</sub> (%)	3.05	2.97	1.13	11.7	0.97	2.50	0.02
Cd	0.67	0.40	0.10	8.60	0.72	3.68	0.54
Sn	13.0	7.10	1.90	462	28	11.84	0.66
Sb	2.66	2.00	0.80	31.1	2.59	5.54	0.49
Cr	74	71.0	24.0	520	27.2	8.91	0.10
Ni	25	22.0	6.00	170	12.2	4.62	0.30
Cu	41	27.0	8.00	446	42.3	4.51	0.61
Zn	131	97.0	35	1576	118.9	5.45	0.52
As	9.75	9.00	2.00	104	5.63	9.56	0.14
Se	0.40	0.40	0.00	2.80	0.27	2.43	0.20
Br	8.66	8.00	2.00	35	3.29	2.24	0.33
Pb	96	64	18	946	94.8	3.93	0.58
U	1.64	1.60	0.00	5.60	0.69	1.34	0.00
LOI(%)	6.30	5.83	1.90	31	2.36	3.18	0.25
pH	5.60	5.62	3.50	7.90	0.95	-0.04	-0.02

## 4.2 CADMIUM CONTAMINATION OF NESTED SURVEY SITES

To determine whether the nested survey data were representative of results from the urban survey, separate summary statistics for the two datasets were also calculated. The most notable difference was in the mean concentration of Cd for the urban ( $0.41 \text{ mg kg}^{-1}$ ) and nested ( $1.34 \text{ mg kg}^{-1}$ ) surveys, which showed that some of the latter sites may have been contaminated. Samples from two of these areas had particularly high Cd contents; mean values of 12 samples from the London Road allotments and Bell Green were 2 and  $2.1 \text{ mg kg}^{-1}$  respectively. In the case of the former, this may be of particular concern if vegetables are grown in this soil and subsequently eaten. The CLEA soil guideline value for total Cd in a soil of pH 6 (mean pH at London Rd is 6.1) is  $1 \text{ mg kg}^{-1}$  (Department of the Environment Food and Rural Affairs and the Environment Agency, 2002). Given that the CLEA model assumes that 99% of the exposure to Cd is through the consumption of homegrown vegetables, it appears there may be a clear source-pathway-receptor linkage for this site. Hence, there may be a potentially significant risk to human health and this is currently being raised with the City Council.

## 4.3 PARTICLE-SIZE DATA

**Table 7 - Average clay, silt and sand content of soils at each of the nested survey centres**

Location and land use	Clay (%)	Silt (%)	Sand (%)
Farm at Longford , Pasture	13	25	62
London Road, Allotments	16	23	60
War Memorial Park, Recreational	17	30	53
Aldermans Green, Nature Reserve	24	40	37
Wilenhall, Playing field / Waste ground	13	18	68
Kennedy school, Playing field	15	32	53
Tilehill Wood School, Playing field	21	34	45
Roman Fort, Pasture	11	23	66
University, Playing field	26	28	46
University, Playing field	17	29	55
Allesley Hall, Park	19	31	51
Bell Green, Waste ground	23	34	43
Binley, Rugby ground/Roadside	23	32	45

The particle-size data can be summarised graphically in the form of a tri-linear diagram, the three axes showing the percentage content of sand, silt and clay sized fractions ( $>60 \mu\text{m}$ ,  $2-60 \mu\text{m}$  and  $<2 \mu\text{m}$  respectively). There was a significant amount of variation in the particle-size distribution *between* the nested survey centres, accounting for 45 % of the variation in clay content. In terms of the UK system for soil classification (Avery and Bascomb, 1982), the samples from the nested survey range from ‘loamy sand’ (the bottom-left hand corner) to ‘clay’ (the centre of Figure 4). There also appeared to be a significant amount of variation within certain nested centres. When classified by parent material type (of which there were thirteen), it accounted for 55% of the variation in clay content.

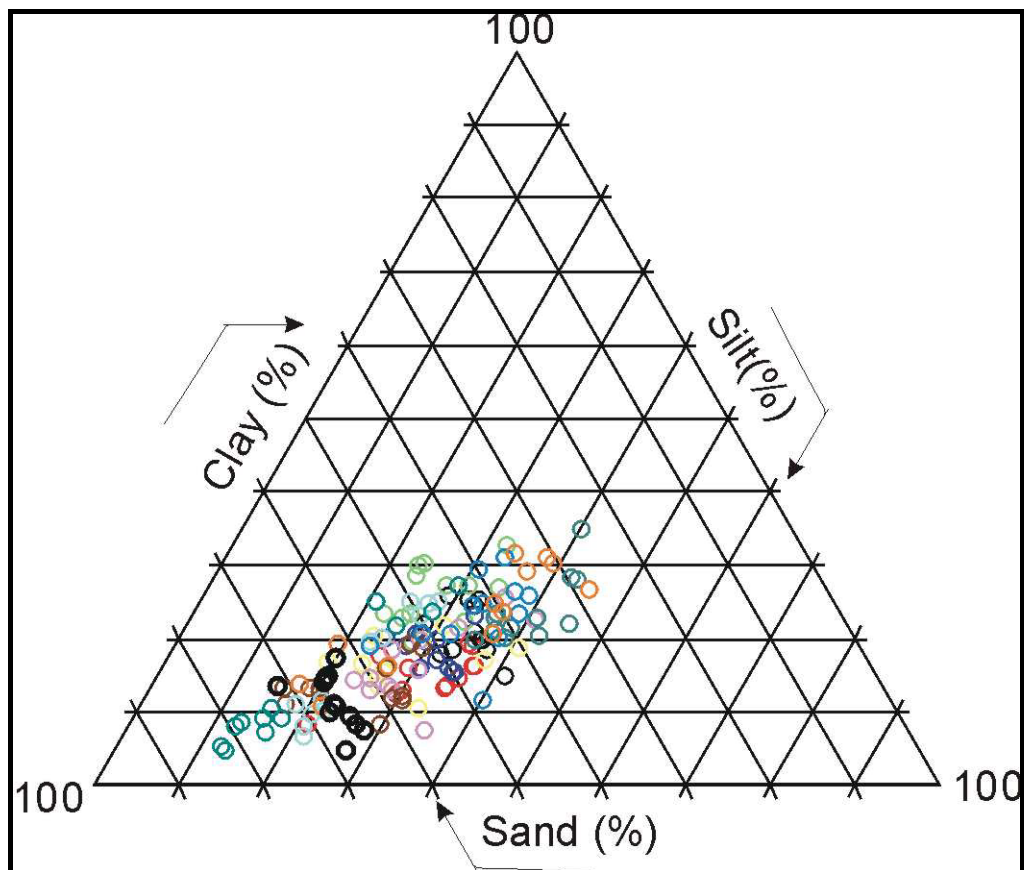


Figure 4 - Particle-size distribution of soils in the nested survey (grouped by colour)

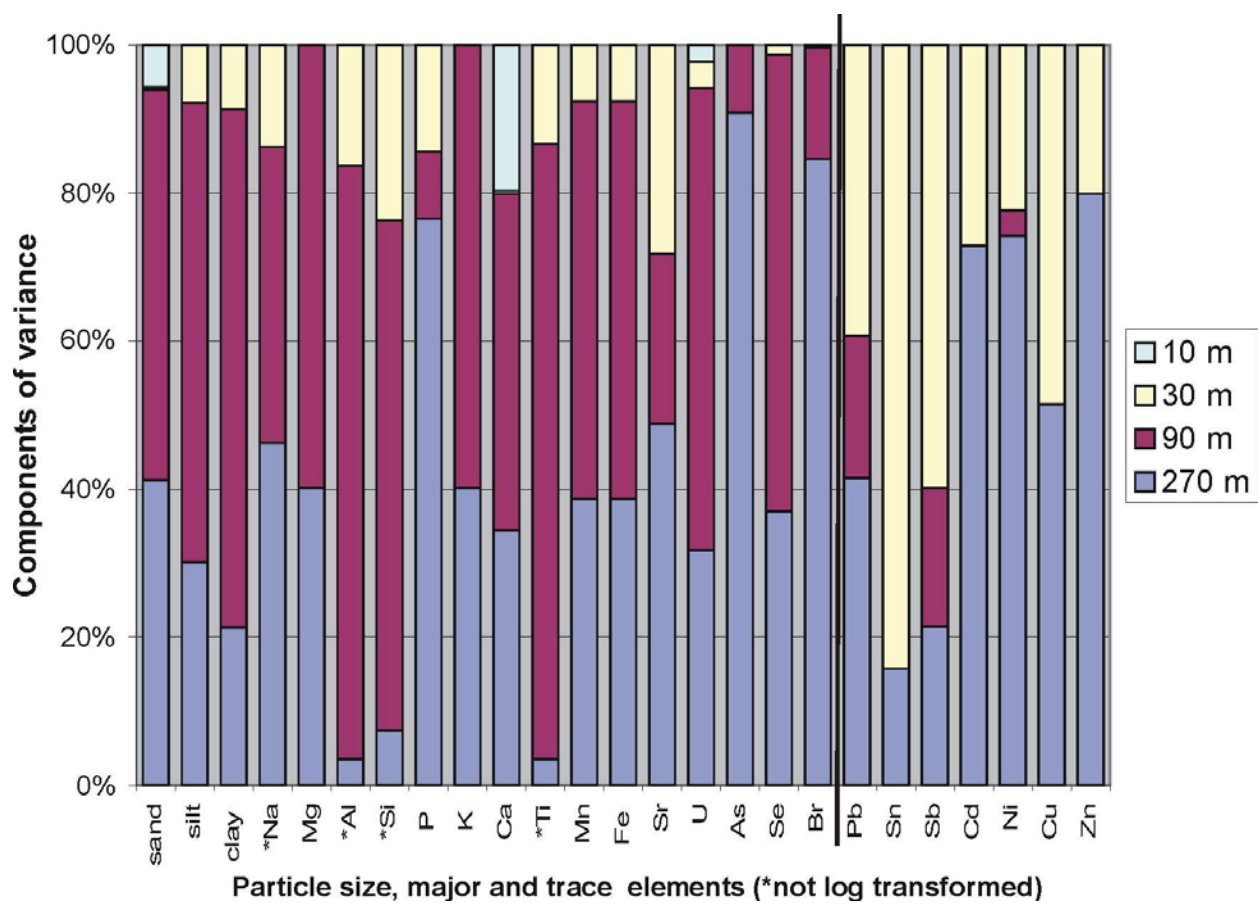


Figure 5 - Components of variation by stage of the nested survey

## 5 Nested analysis of variance

### 5.1 COMPONENTS OF VARIANCE

Using the method described in Table 2 and a Fortran computer programme written by Dr Richard Webster, the components of variance at each of the four spatial scales were calculated. The results are presented in Figure 5. Based on our experience of natural soil variation, we would expect to find greater components of variance over larger spatial scales. For the vast majority of the elements this appears to be the case for the nested samples (see Figure 5). Up to 80% of the variance for particle size parameters, and for a series of major and trace elements to the left of the vertical black line was accounted for in the 90 and 270 metre scales. However, in the case of Pb, Sn, Sb, Cd, Ni, Cu and Zn, a larger proportion (22-75%) of the variation occurred at spatial scales of 30 metres or less. Without further information it is not possible to determine whether this variation is due to anthropogenic impacts (pollution) or natural variation. However, many of these elements are common environmental pollutants suggesting that their greater short-scale variability may result from human activities.

## 6 Geostatistical analysis

Rather than attempting to undertake analysis of every element, it was necessary to select around five which could be used for assessing geostatistical methods. As discussed above, and more comprehensively by Lark (2000), robust methods tend to be effective when applied to strongly skewed distributions (Table 6) which have a significant number of outliers. It was decided to focus on those elements which may be of most serious concern to human health. Therefore Cd, As, Pb, Zn and Ni were selected for more detailed analysis. The methods described in sections 1.2 were applied to assess whether classical or robust methods would be more appropriate variogram estimators for these five elements.

### 6.1 DATA TRANSFORMATION

Descriptive statistics were calculated to determine whether data transformation was necessary (Table 8). In each case the skew coefficient was greater than 1, and in most cases the octile skew was larger than 0.2, suggesting that transformation was necessary prior to further analysis. In each case, transformation to natural logarithms reduced the skewness coefficients to acceptable values, and these data were used for variogram estimation.

**Table 8 - Descriptive statistics for the total concentration of five elements in urban and nested soil surveys (n=551) used in the geostatistical analysis and the skewness co-efficients of the log<sub>e</sub> transformed data.**

	<b>Cd</b>	<b>As</b>	<b>Pb</b>	<b>Zn</b>	<b>Ni</b>
<b>Mean (mg kg<sup>-1</sup>)</b>	0.7	9.8	95.6	131	24.6
<b>Median (mg kg<sup>-1</sup>)</b>	0.4	9	64	97	22
<b>St. Dev. (mg kg<sup>-1</sup>)</b>	0.7	5.6	94.8	118	12.2
<b>Skew</b>	3.7	9.6	3.9	5.5	4.6
<b>Number outliers (&gt;mean+3 S.D.)</b>	6	4	2	11	5
<b>Octile skew</b>	0.54	0.14	0.58	0.52	0.30
<b>Log<sub>e</sub> skew</b>	0.02	0.7	0.8	1.1	0.4
<b>Log<sub>e</sub> octile skew</b>	-0.05	-0.05	0.25	0.25	0.10

## 6.2 ASSESSING THE PERFORMANCE OF VARIOGRAM ESTIMATORS

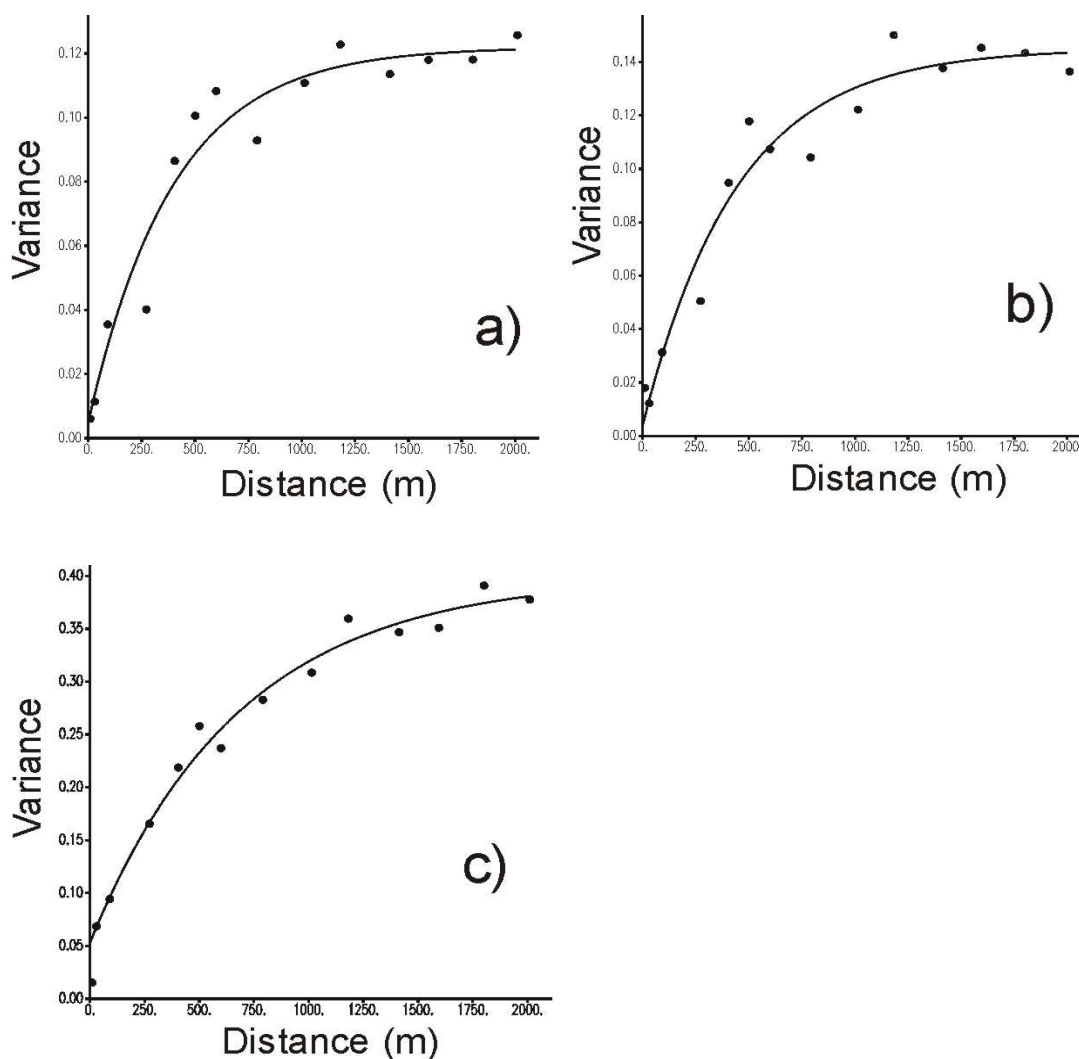
Experimental values for the variogram were calculated at pre-determined lag intervals between 10 and 2000 metres for each of the five variates using the Matheron (Ma), Cressie-Hawkins (CH) and Dowd estimators. In each case authorised models were fitted to the variograms using a weighted least-squares procedure. The median  $\theta$  statistic was determined to assess the performance of the various models by comparison with its proximity to the theoretical value of 0.455 and the bootstrapped confidence interval of between 0.37 and 0.55. The results of the variogram modelling and cross validation exercise are shown in Table 9.

**Table 9 – Parameters of the variogram models and the cross validation assessment.**

Element	Estimator	Mod*	Co†	C1‡	** A1 (m)	Median theta
Cd	Ma	Sp	0.103	0.977	681.7	0.406
As	Ma	Exp	0.005	0.117	1193.1	0.403
Pb	Ma	Exp	0.053	0.346	2049	0.396
Zn	Ma	Exp	0.049	0.266	2304	0.275
	CH	Sp	0.031	0.237	1766	0.413
	Dowd	Sp	0	0.241	1700	0.935
Ni	Ma	Exp	0.004	0.141	1326.9	0.444

Model abbreviations: Sp (spherical) Exp (Exponential)

† - Nugget variance; ‡ - Correlated variance; \*\* range of the model in metres

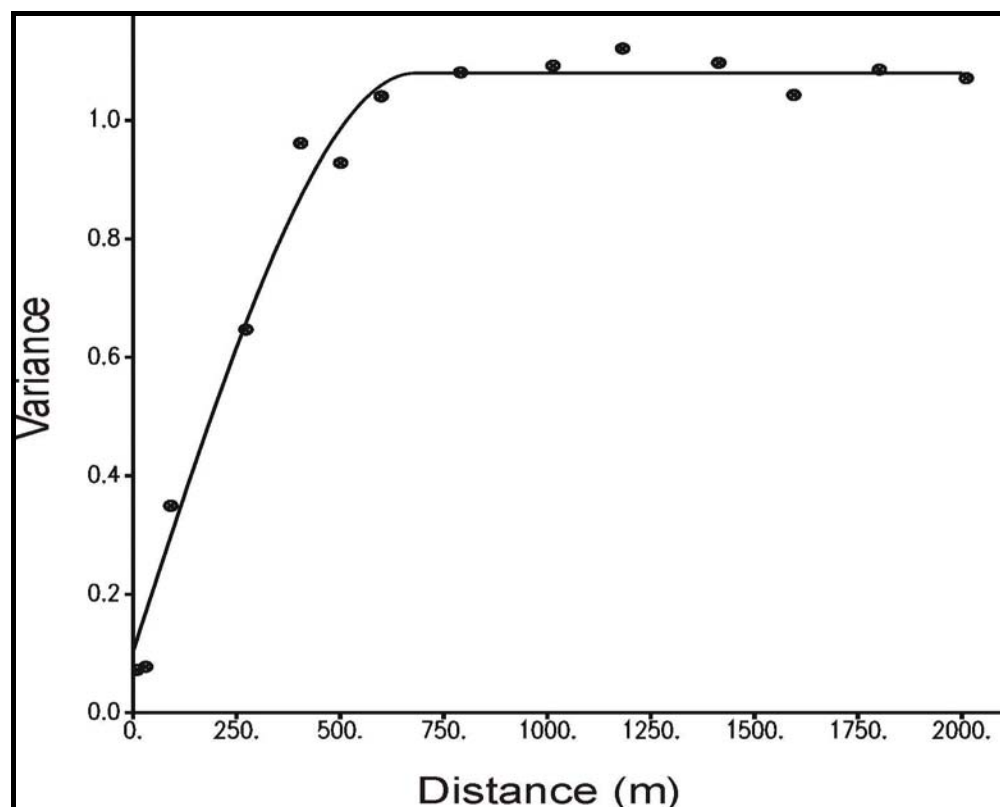


**Figure 6 - Experimental variograms (symbols) and models fitted to them (lines) for  $\log_e$  transformed data.**

Parameters for the models are given in Table 9.

Based on the median theta values calculated from the cross-validation exercise, Matheron's estimator of the variogram was found to be optimal for the estimation of Cd, As, Pb and Ni (see Figure 6 and Figure 7). However, the Cressie-Hawkins robust estimator was best suited for estimation of Zn (Figure 10). Therefore, for the purpose of estimating values of Cd, As, Pb and Ni at unsampled locations in Coventry, the use of conventional geostatistics is appropriate (see section 6.3 for an example). Given that these elements were selected because of their significantly skewed distributions (where robust geostatistics were likely to be more appropriate), it is likely that Matheron's estimator would also be selected for the majority of the other elements which we did not analyse in more detail. Although these data all have skewed distributions, they have few outlying values (**Table 8**) and it appears that the number of outliers may be significant in determining whether conventional or robust methods are optimal. For example, in the case of Zn, there were significantly more outlying values (11 as opposed to between 2 and 6), and the Cressie-Hawkins robust estimator was optimal. In section 6.4 the spatial outliers in the Zn dataset are identified and in section 6.5 we present kriged isarithmic maps of the data, with and without the spatial outliers.

### 6.3 KRIGING CADMIUM WITH MATHERON'S STANDARD ESTIMATOR



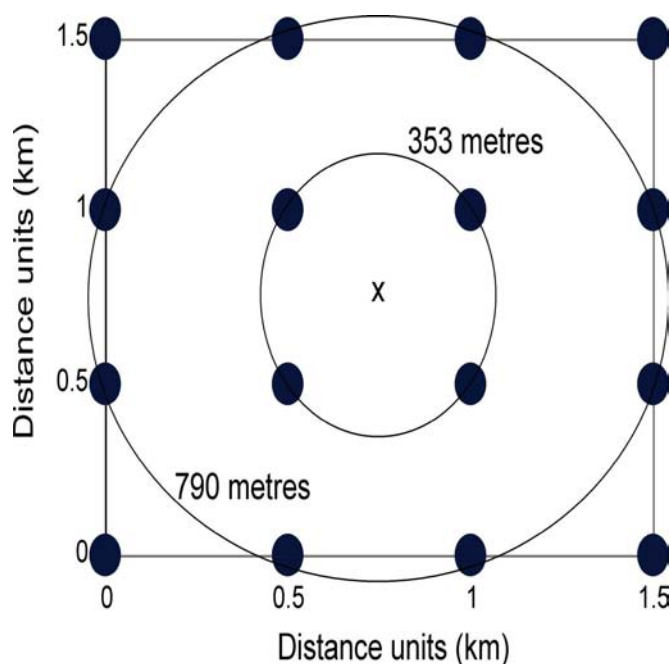
**Figure 7 - Isotropic spherical model (solid line) fitted to the experimental variogram (circles) using Matheron's estimator -  $\log_e$  transformed Cd data.**

Table 9 lists the model parameters.

Analysis of the results from the cross-validation exercise showed that Matheron's standard estimator of the variogram was optimal for Cd. A weighted least-squares procedure in the programme Genstat (Payne et al., 1996) was used to select a spherical model, which was then fitted to the estimates (Figure 7). It is important to note at this stage that the range of the model (the limit of spatial autocorrelation for Cd) of 680 metres, shows that based on the typical sampling density adopted in urban surveys, only the four nearest sample points (surrounding a point at which a value is to be estimated) will have significant weights used for estimation in the kriging model (illustrated in Figure 8). Sample points beyond this range are as likely (on



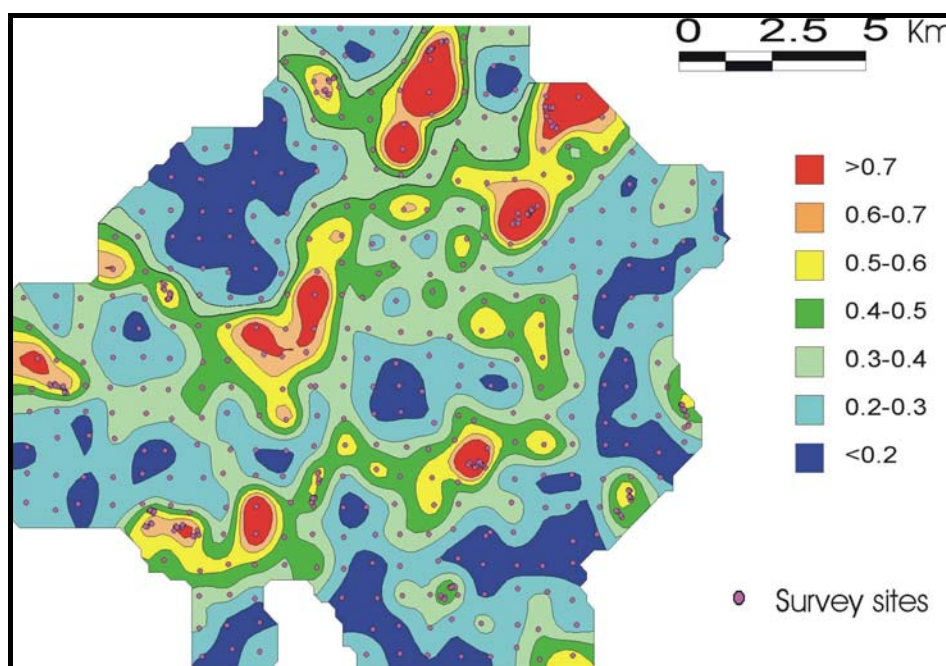
average) to be as different from all other points in the survey, to the point at which the value is being estimated.



**Figure 8 – Radii of circles around an unsampled location (x) where a value is to be estimated in relation to actual sample points (•) on a 500 metre\* sampling interval.**

\* Although the sample sites in the urban survey are not collected on a precise 500m grid, the average distance to neighbouring points across the survey area will approximate to a regular grid as shown.

After a natural logarithm transform, ordinary kriging was used to estimate values on a regular grid of points separated by 200 metres. The values were then backtransformed and contours threaded through them using concentration intervals of  $0.1 \text{ mg kg}^{-1}$  (Figure 9). The map identifies considerable spatial variation in the Cd content throughout Coventry, with several discrete areas of elevated values ( $>0.7 \text{ mg kg}^{-1}$ ), some of which coincide with the nested survey centres.



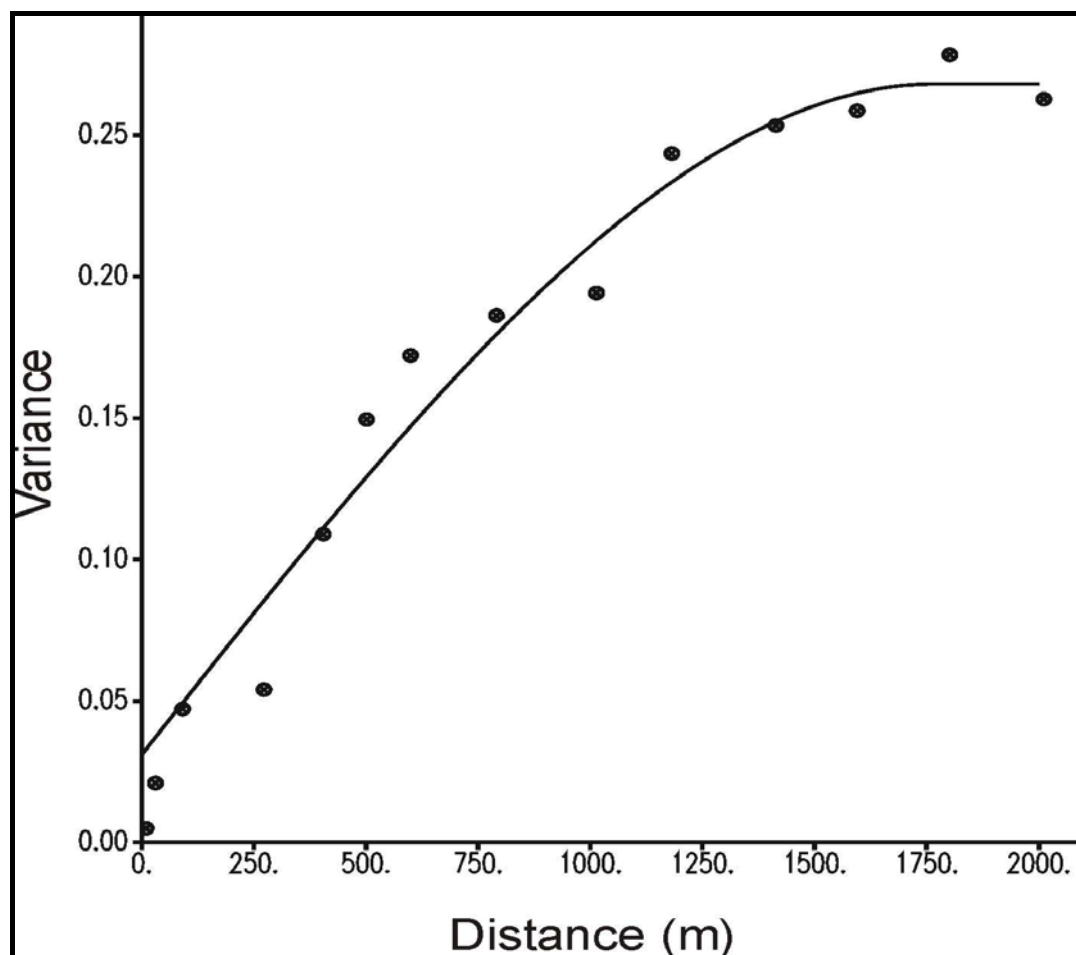
**Figure 9 - Kriged map of Cd content ( $\text{mg kg}^{-1}$ ) of top-soils throughout Coventry using Matheron's standard estimator**

## 6.4 IDENTIFICATION OF SPATIAL OUTLIERS (ZINC DATA)

In section 1.3, a brief description of a method for the identification of spatial outliers was provided, summarised from Lark (2002). This methodology was applied to the data for Zn using the Cressie-Hawkins robust estimator of the variogram. The threshold for values of  $O(x)$  which indicate the sample may be part of a contaminating process was calculated to be 2.58. Of the values of  $O(x)$ , 29 were larger than this threshold and were therefore identified as representing the contaminant process. These spatial outliers are referred to in the subsequent sections of the report.

## 6.5 KRIGING TOPSOIL ZINC WITH AND WITHOUT SPATIAL OUTLIERS

Of the five trace elements studied, the data for Zn was found to require the use of a robust estimator of the variogram. Using the Cressie-Hawkins estimator, experimental values of the variogram were calculated at a range of lag intervals. Using a weighted least-squares procedure, a spherical model was selected and this was fitted (Figure 10). In the case of Zn, the range of autocorrelation is 1700 metres, far larger than for Cd. A comparison of the two variograms shows that variances increase more gradually with distance for Zn than for Cd. Hence, estimation of Zn at unsampled locations will be based on significantly more sample points than was the case for Cd (the 1700 metre range for Zn is beyond the outer circle shown in Figure 8).

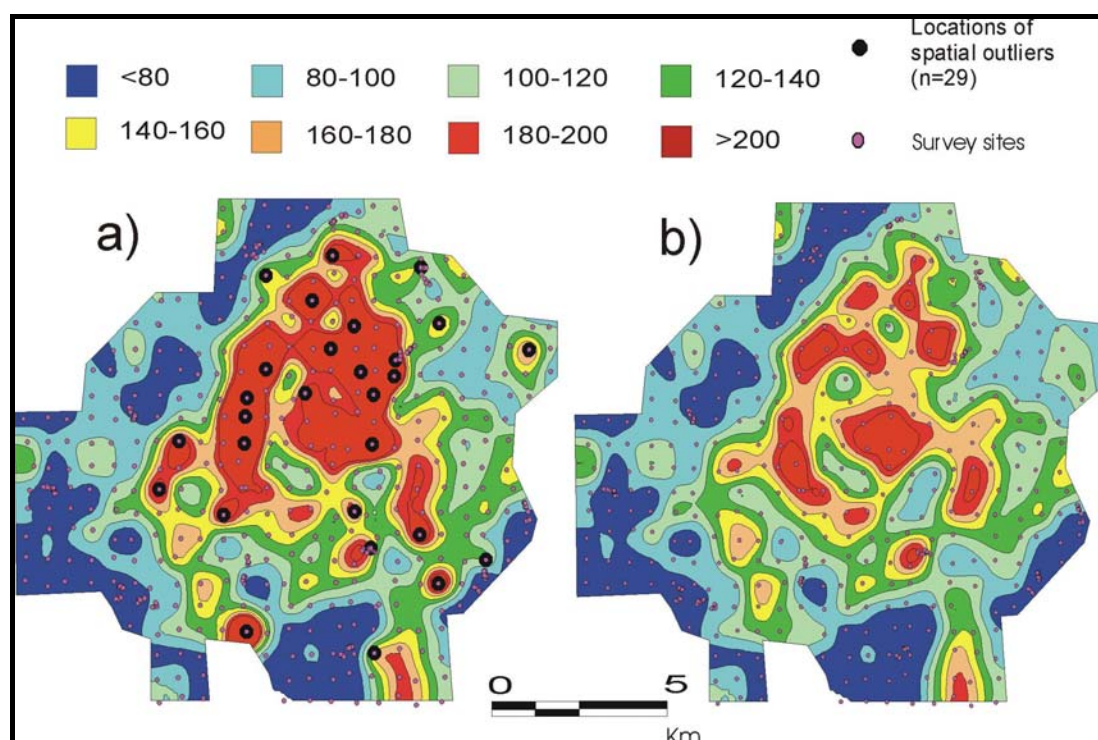


**Figure 10 - Isotropic spherical model (solid line) fitted to experimental variogram (circles) using the Cressie-Hawkins robust estimator:  $\log_e$  transformed Zn data.**

Model parameters in Table 9.

Following the natural log transformation, ordinary kriging was used to estimate values on a regular grid of points separated by 200 metres, and these values were backtransformed. Contours were threaded through the estimated values using concentration intervals of  $20 \text{ mg kg}^{-1}$

(Figure 11a). The locations of the 29 spatial outliers were also highlighted. In addition, the same procedure was undertaken after removal of the 29 spatial outliers to determine the scale of the impact of the point-process (Figure 11b).



**Figure 11 - Kriged maps showing Zn content ( $\text{mg kg}^{-1}$ ) of top-soils in Coventry : a) with, and b) without, spatial outliers**

The highest top-soil Zn concentrations (representing the background process) are found in the centre and to the north of the City, where there is also a high frequency of spatial outliers (representing a point/contaminant process). By removing the spatial outliers, there is a significant change in the distribution in the region of the highest values (see Figure 11), the two highest concentration classes cover a smaller area. In addition, a number of the ‘bullseye’ patterns which were associated with spatial outliers in Figure 11a do not appear in Figure 11b. The latter map would be more suited to the estimation of background values at specific sites throughout the city. More detailed fieldwork could be undertaken around the location of the spatial outliers to determine the extent of the contaminating process.

## 7 Discussion and Conclusions

### 7.1 USE OF ROBUST VERSUS CLASSICAL GEOSTATISTICS

Matheron’s standard estimator of the variogram was optimal for four of the five trace elements studied. As these elements had the most skewed distributions, it is likely that the optimum estimator for the other major and trace elements would also be that of Matheron. In the case of Zn, a robust estimator performed better than the standard estimator. This may be related to the large number of outliers in the Zn dataset (compared to the other trace elements). It is therefore recommended that when a large number of outliers are present, a comparison of robust and classical estimators is undertaken (using the cross validation procedure described by Lark (2002) to determine whether a robust estimator should be used. *To determine whether robust geostatistical methods should be applied to other GSUE data, there is a need to determine the skewness coefficients and the number of outliers in these datasets. We will then have a better*

*understanding of whether the statistical distributions of the Coventry data are typical of those in the other centres.*

## **7.2 THE OBJECTIVES OF UNDERTAKING GEOCHEMICAL SURVEYS IN THE URBAN ENVIRONMENT**

The overriding objective of the urban geochemical surveys undertaken by the British Geological Survey was to complement the broader, rural geochemical baseline of the UK landmass. The main applications of these two surveys are quite different. The principal use of soil geochemical data in the urban environment is related to environmental protection and contaminated land regulations, undertaken by local authorities. Urban soil geochemical survey data can be used to establish the typical concentrations of contaminants in certain areas of a town or city (typically somewhat higher than over adjacent parent materials in the rural environment) or, identify more significant contamination (of unknown spatial extent) at individual sites. In the case of the former, it is often desirable for the user to have a continuous surface map showing variation across the whole area. The user may then wish to interrogate point data for a specific area in more detail. For locations where individual (composite) samples indicate the presence of significant contamination, more detailed surveys are required to establish the precise extent of the contamination. This raises the question: what is an appropriate sampling resolution for user requirements given the perceived need for continuous surface maps, bearing in mind the increased cost associated with more detailed sampling?

## **7.3 SAMPLING INTENSITY VERSUS ESTIMATION UNCERTAINTY**

It has been demonstrated that there is spatial correlation in soil geochemical properties in Coventry at the chosen sampling intensity. As sampling intensity (or resolution) increases, the uncertainty associated with estimating values at unsampled locations (using geostatistical methods) decreases. Selection of a sampling resolution for subsequent urban surveys to meet a pre-determined average estimation uncertainty for a specific element could be attempted. However, the aim of a GSUE survey is not to map specific parameters, but a range of major and trace elements. The spatial variability of the different contaminant elements will vary both within an individual urban area (see the ranges in (see Table 6) and also between urban areas. In addition, the analytical methodology (XRFS) permits the analysis of up to 40 major and trace elements (almost simultaneously). It would be inefficient to collect samples at different resolutions for the analysis of selected elements. Therefore a sampling resolution needs to be selected which is sufficiently dense to ensure that a certain proportion of the spatially correlated variance for the majority of the contaminant-type elements is captured. *Geostatistical analysis of urban geochemical data has indicated to date that the current sampling resolution (4 samples per square kilometre) does capture a certain proportion of the spatially correlated variation. However, when producing a continuous surface map, it is essential first to undertake exploratory data analysis and construct variograms of the data to assess the degree of autocorrelation prior to kriging. If little or no autocorrelation is present, it is preferable to present the data as proportional symbols because there is little or nothing to be gained from interpolation.*

When a continuous surface map has been produced it should be accompanied by a description of the method used to create it, and a statement that the contour intervals represent estimates, not true values.

## **7.4 TARGETED SAMPLING**

To increase the utility of the soil geochemical data collected in the urban environment, it would be beneficial to focus on perhaps two or three key sites in the city where human exposure to contamination may be significant, such as allotments. These sites could be selected in conjunction with the City/Local Council on the basis of land use information which may indicate the likely presence of historical contamination.

# Glossary

**Autocorrelation** – the technical term to describe the nature of variation in environmental variables whereby samples closer together tend to have more similar properties than those further apart.

**Cross Validation** - A technique for testing the validity of a variogram model by kriging each sampled location with all of the other samples in the search neighborhood, and comparing the estimates with the true sample values. Interpretation of results, however, can often be difficult. Unusually large differences between estimated and true values may indicate the presence of “spatial outliers”, or points which do not seem to belong with their surroundings.

**Geostatistics** - A methodology for the analysis of spatially correlated data. The characteristic feature is the use of variograms or related techniques to quantify and model the spatial correlation structure. Also includes the various techniques such as kriging, which utilize spatial correlation models.

**Matheron’s estimator (of the variogram)**- one-half the mean squared difference of paired sample measurements as a function of the distance (and optionally of the direction) between samples.

**Kriging** – a weighted-moving-average interpolation method where the set of weights assigned to samples minimizes the estimation variance, which is computed as a function of the variogram model and locations of the samples relative to each other, and to the point or block being estimated

**Lag** - a distance class interval used for variogram computation.

**Linear Model** - a function frequently used when fitting mathematical models to experimental variograms, often in combination with a nugget model.

**Ordinary Kriging** - a variety of kriging which assumes that local means are not necessarily closely related to the population mean, and which therefore uses only the samples in the local neighborhood for the estimate. Ordinary kriging is the most commonly used method for environmental situations.

**Range** - the distance at which the model fitted to experimental values of the variogram reaches its maximum value, or sill.

**Robust estimator** (of the variogram) – formulas for calculating robust estimates of variance with increasing lag distance. Examples include that of Cressie and Hawkins (based on the mean of the absolute squared difference between sample pairs) and Dowd (based on the median absolute deviation).

**Variogram** – a plot of the variance (one-half the mean squared difference) of paired sample measurements as a function of the distance (and optionally of the direction) between samples. Typically, all possible sample pairs are examined, and grouped into classes (lags) of approximately equal distance and direction. Variograms provide a means of quantifying the commonly observed relationship that samples close together will tend to have more similar values than samples far apart.

**$z(x_i)$**  – value of the parameter ( $z$ ) measured at location  $x$ , for which there are  $i=1, 2, 3, \dots, n$  values in a dataset

$$\hat{Z}(x_i)$$

- estimated value at a location, based on the removal of the original value at that location from the dataset, and its estimation based on values at adjacent locations using ordinary kriging

$$\hat{\sigma}_K(x_i)$$

- kriging variance at the location where the value  $Z(x_i)$  is estimated through cross-validation.

# References

- Avery, B W, and Bascomb, C L. 1982. Soil survey laboratory methods. *Soil survey Technical Monograph*, No. 6.
- Burrough, P A, and McDonnell, R A. 2000. *Principles of Geographical Information Systems*. (Oxford University Press).
- Cressie, N, and Hawkins, D. 1980. Robust estimation of the variogram. *Mathematical Geology*, Vol. 12, 115-125.
- Department of the Environment Food and Rural Affairs and the Environment Agency. 2002. Soil Guideline Values for Cadmium Contamination.
- Dowd, P A. 1984. The variogram and kriging: robust and resistant estimators. 91-106 in *Geostatistics for Natural Resources Characterization*. Verly, G, David, M, Journel, A G and Marechal, A (Eds.), Dordrecht, Reidel.
- Fordyce, F M and Ander, E. L. 2003. Urban Geochemistry GIS - A Case Study from Stoke-on-Trent. *British Geological Survey Internal Report IR/01/35R*.
- Ingham, M N, and Vrebis, B A R. 1994. High Productivity Geochemical XRF Analysis. *Advances in X-ray Analysis*, Vol. 37, 717-724.
- Lark, R M. 2000. A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, Vol. 51 (1), 137-157.
- Lark, R M. 2002. Modelling complex soil properties as contaminated regionalized variables. *Geoderma*, Vol. 106 (3-4), 173-190.
- Oliver, M A, and Webster, R. 1986. Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. *Geographical Analysis*, Vol. 18 (3), 227-242.
- Payne, R W, Webster, R, and Mugglestone, M A. 1996. An overview of spatial analysis in Genstat. *GENSTAT96: a statistical conference for GENSTAT users, Waite Agricultural Research Institute, The University of Adelaide, South Australia, December 4-6, 1996*, pp. 27.
- Webster, R, and Oliver, M A. 2001. *Geostatistics for Environmental Scientists*. Statistics in Practice (Chichester: John Wiley & Sons).
- Youden, W J, and Mehlich, A. 1937. Selection of efficient methods for soil sampling. *Contributions of the Boyce Thompson Institute of Plant Research*, Vol. 9, 59-70.